WavLLM: Towards Robust and Adaptive Speech Large Language Model

Year: 2024 | *Citations: 104* | *Authors: Shujie Hu, Long Zhou, Shujie Liu*

Abstract

The recent advancements in large language models (LLMs) have revolutionized the field of natural language processing, progressively broadening their scope to multimodal perception and generation. However, effectively integrating listening capabilities into LLMs poses significant challenges, particularly with respect to generalizing across varied contexts and executing complex auditory tasks. In this work, we introduce WavLLM, a robust and adaptive speech large language model with dual encoders, and a prompt-aware LoRA weight adapter, optimized by a two-stage curriculum learning approach. Leveraging dual encoders, we decouple different types of speech information, utilizing a Whisper encoder to process the semantic content of speech, and a WavLM encoder to capture the unique characteristics of the speaker's identity. Within the curriculum learning framework, WavLLM first builds its foundational capabilities by optimizing on mixed elementary single tasks, followed by advanced multi-task training on more complex tasks such as combinations of the elementary tasks. To enhance the flexibility and adherence to different tasks and instructions, a prompt-aware LoRA weight adapter is introduced in the second advanced multi-task training stage. We validate the proposed model on universal speech benchmarks including tasks such as ASR, ST, SV, ER, and also apply it to specialized datasets like Gaokao English listening comprehension set for SQA, and speech Chain-of-Thought (CoT) evaluation set. Experiments demonstrate that the proposed model achieves state-of-the-art performance across a range of speech tasks on the same model size, exhibiting robust generalization capabilities in executing complex tasks using CoT approach. Furthermore, our model successfully completes Gaokao tasks without specialized training. The codes, models, audio, and Gaokao evaluation set can be accessed at \url{aka.ms/wavllm}.